Introduction to Clinical Trials - Day 2

Extra - Survival and Change from Baseline Endpoints

Presented July 26, 2016

Susanne J. May
Department of Biostatistics
University of Washington

Daniel L. Gillen
Department of Statistics
University of California, Irvine

©2016 Daniel L. Gillen, PhD and Susanne J. May, PhD

Time-to-event outcomes

Properties of censored time-to-event data

- Time-to-event data is common in the health sciences; for example:
  - Prolong survival (delay death)
  - Prolong remission time (delay recurrence)
  - Prevent MI (delay time until MI)
  - Prevent cancer (delay time until cancer is detected)
  - Reduce time until discharge from hospital
  - Prolong time between hospitalizations
- A feature of this type of data:
  - We know the time of the event for some subjects.
  - For other subjects we only know the amount of time without the event.

Time-to-event outcomes

Properties of censored time-to-event data

- Time to event outcomes are bivariate (2 variables in one):
  - Time
  - Indicator for presence of the event (Y/N)
- Censoring:
  - A “censored observation" is an individual who has not had the event.
  - A censored observation is an example of “non-ignorable" missing data.
- Classic (right) censoring mechanism:
  - Subjects enter a study at different times so at the time of analysis there is a different amount of follow-up on each individual.
  - We know only that the event has not occurred before time T.
- Appropriate statistical methods (survival analysis) must be used to account for the censoring.

Time-to-event outcomes

Properties of censored time-to-event data

- Example: Survival following stroke in patients with coronary artery disease (ordered by survival time).

Patient Number	Died (1 = Yes)	Weeks Since Stroke
14	1	0.4
3	0	0.7
15	1	1.1
13	1	20.8
7	0	31.7
4	0	35.9
9	0	43.3
8	0	55.5
6	0	70.9
5	0	76.6
11	1	78.1
12	1	78.4
2	0	94.7
1	0	165.6
10	0	168.8

Properties of censored time-to-event data

Estimating Survival Probability

- Incorrect estimates:
  - Throw out all missing (censored) data:
    52-week survival: 2/5 = 0.4
  - Throw out only those censored before the time point:
    52-week survival: 8/11 = 0.73
- Inefficient estimates:
  - Throw out all subjects who have not been in the study for the time of interest.
- Correct approach:
  - Kaplan-Meier (product limit) estimates:

    Consistent estimate of the percent with the event as a function of follow-up time.

Properties of censored time-to-event data

SISCR
UW - 2016

Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline

Estimating Survival Probability

Example: K-M curve for progression-free survival (OCEANS trial)

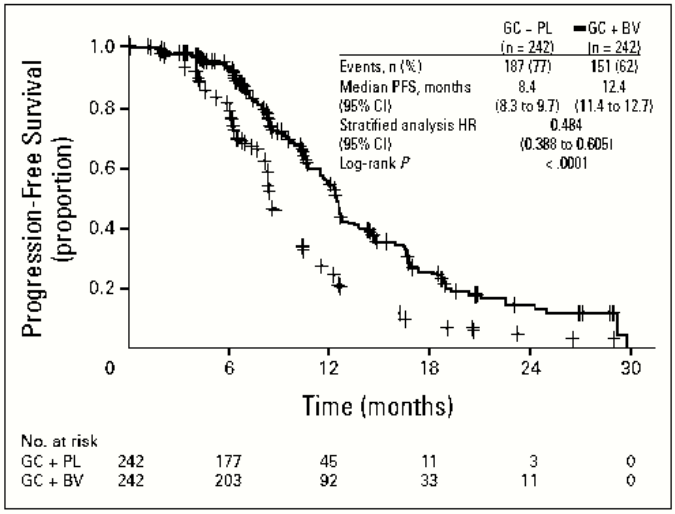


Fig 2. Kaplan-Meier estimates of progression-free survival (PFS) based on investigator assessment, censoring for non–protocol-specified therapy (randomly assigned patients). BV, bevacizumab; GC, gemcitabine plus carboplatin; HR, hazard ratio; PL, placebo.

Properties of censored time-to-event data

Summarizing the K-M curve

- The K-M curve is an estimate of the distribution of the individual data (like a probability distribution).
- We need to select a functional and contrast to measure treatment effects.
  - Functional: characteristic of the distribution to summarize the outcome in the population.
  - Contrast: how to measure differences between outcomes in two populations.

Properties of censored time-to-event data

SISCR
UW - 2016

Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline

Summarizing the K-M curve

- Common choices for functional and contrast:
  - Risk difference at an index time; for example:
    θ_0 = risk of event before 12-months with placebo.
    θ_1 = risk of event before 12-months with treatment.
    $\theta = \theta_1 - \theta_0$: difference in risk of event by 12-months.
  - Ratio of Poisson incidence; for example:
    θ_0 = death rate without screening.
    θ_1 = death rate with screening.
    $\theta = \frac{\theta_1}{\theta_0}$: rate ratio.
  - Hazard ratio:
    θ_0 = hazard of progression with placebo.
    θ_1 = hazard of progression with treatment.
    $\theta = \frac{\theta_1}{\theta_0}$: hazard ratio.
  - Restricted mean survival (area under the K-M curve):
    θ_0 = Mean years lived without screening.
    θ_1 = Mean years lived in with screening.
    $\theta = \theta_1 - \theta_0$: average number of years of live saved (over 5 years).

Properties of censored time-to-event data

Summarizing the K-M curve: PLCO Example

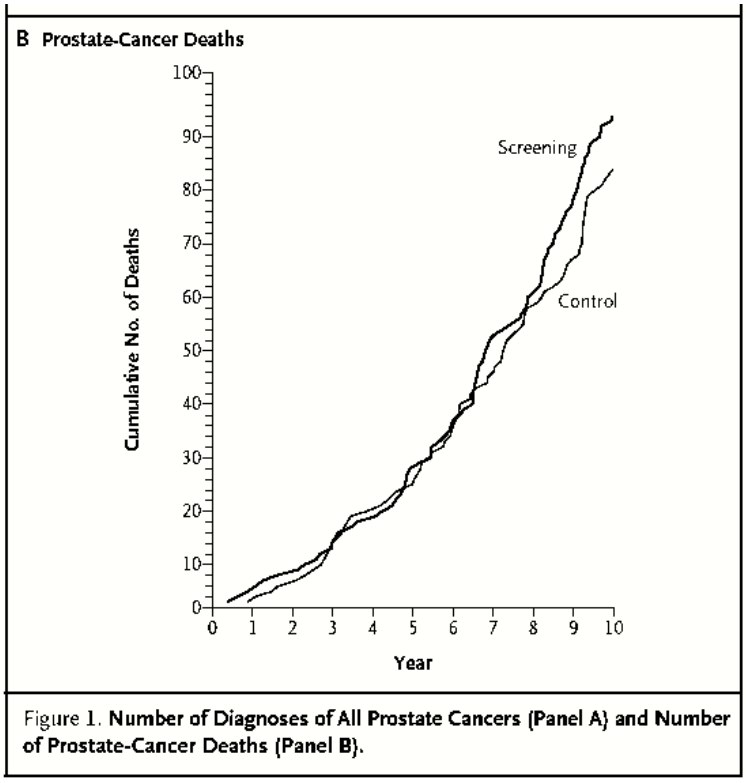


Figure 1. **Number of Diagnoses of All Prostate Cancers (Panel A) and Number of Prostate-Cancer Deaths (Panel B).**

Properties of censored time-to-event data

SISCR
UW - 2016

Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline

Summarizing the K-M curve: PLCO Example

- Some possible choices for a functional and contrast:
  - Difference in 8-year mortality risk
  - Ratio of incidence rates (deaths per person-year)
  - Ratio of hazards
  - Mean years of life saved (over 8 years)

Properties of censored time-to-event data

Summarizing the K-M curve: PLCO Example

- Information using Poisson parameterization:

$$D = \left(\frac{2 \times 1.96}{log(0.5)}\right)^2 \times 2 = 64$$

  i.e., 64 deaths per group or 128 deaths total.

- Information using hazard ratio parameterization (note D is the TOTAL number of deaths:

$$D = \left(\frac{2 \times 1.96}{log(0.5)}\right)^2 \times 4 = 128$$

  [A poisson probability model is one particular type of proportional hazards models.]

- It is useful to compare properties of different approaches to parameterizing survival (time-to-event) data.

Properties of censored time-to-event data

Common choices for functional and contrast

- Let's consider five different approaches to analyzing survival data. Five statistical models:
  - Model A: Semi-parametric (hazard ratio) model
  - Model B: Fully-parametric (Poisson rate ratio) model
  - Model C: Non-parametric (restricted mean) model
  - Model D: Non-parametric (index time) model
- Simulation can be used to demonstrate how these approaches behave under different true probability models:
  - Truth 1: Exponential failure times
  - Truth 2: Proportional hazards
  - Truth 3: Non-proportional hazards
- Remember: we never know the true probability model, so whichever approach that we choose needs to behave well under any true probability model.

Properties of censored time-to-event data

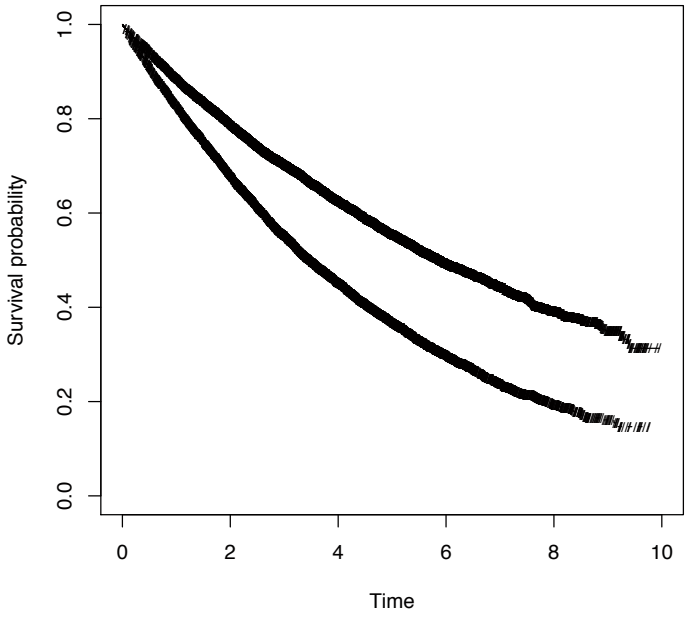
Evaluating the common choices for functional and contrast

- Evaluating all combinations of:
  - Choice of functional and contrast:
    - Model A: Semi-parametric (hazard ratio) model
    - Model B: Fully-parametric (Poisson rate ratio) model
    - Model C: Non-parametric (restricted mean) model
    - Model D: Non-parametric (index time) model
  - True probability distribution:
    - Truth 1: Exponential failure times
    - Truth 2: Proportional hazards
    - Truth 3: Non-proportional hazards
  - Nature of follow-up (type of censoring):
    - Scenario 1: Early censoring
    - Scenario 2: Mid censoring
    - Scenario 3: Late censoring
    - Scenario 4: No censoring

Evaluating the common choices for functional and contrast

Truth 1: Exponential failure times

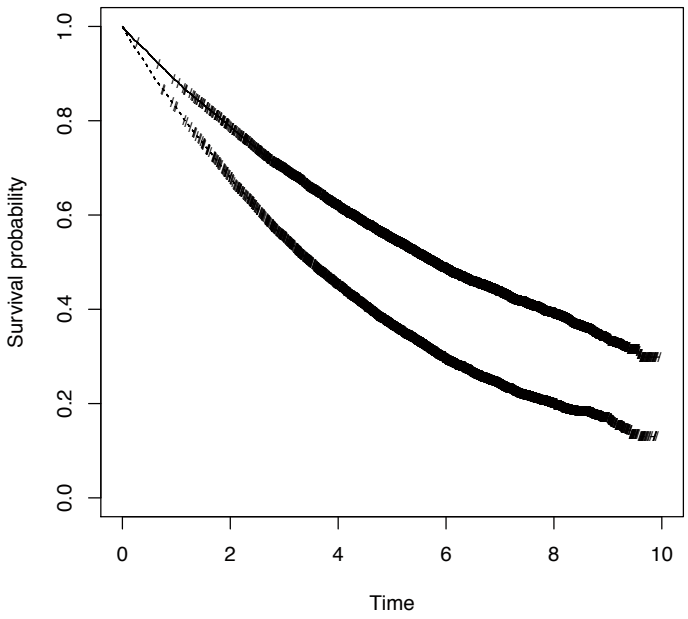
Simulation scenario: early Censoring



Evaluating the common choices for functional and contrast

Truth 1: Exponential failure times

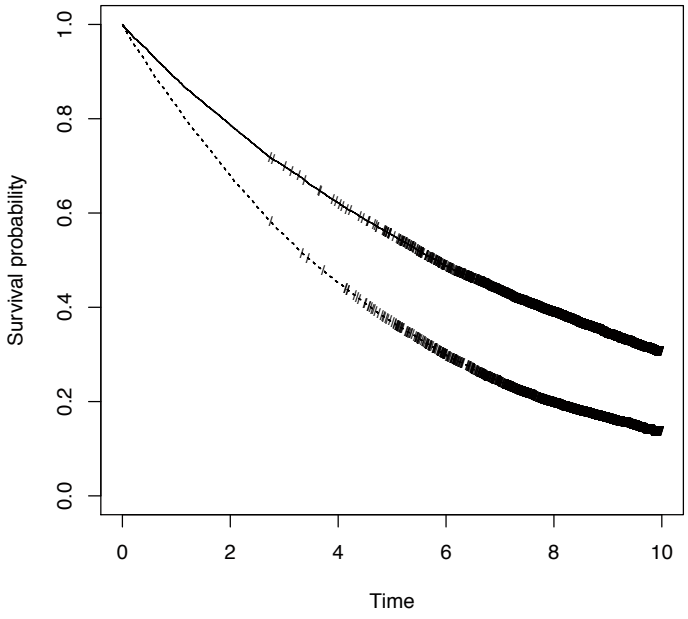
Simulation scenarios: Mid Censoring



Evaluating the common choices for functional and contrast

Truth 1: Exponential failure times

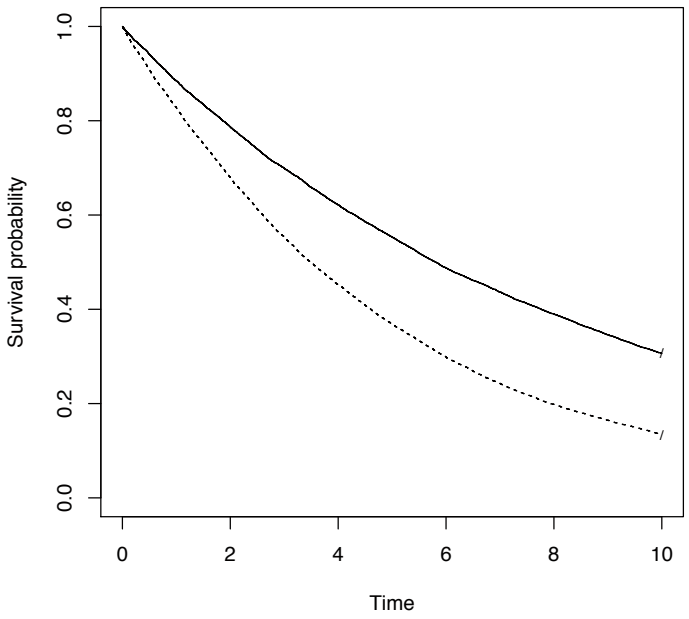
Simulation scenario: Late Censoring



Evaluating the common choices for functional and contrast

Truth 1: Exponential failure times

Simulation scenario: Censoring at 10 only (“No censoring")



Evaluating the common choices for functional and contrast

Truth 1: Exponential failure times

- Results from different parameterizations:

Scenario	Hazard Ratio	Rate Ratio	Restricted Mean	6-year Survival*
Early censoring	0.590	0.591	1.558	0.194
Mid censoring	0.596	0.597	1.504	0.191
Late censoring	0.594	0.594	1.512	0.188
No censoring	0.593	0.594	1.513	0.188

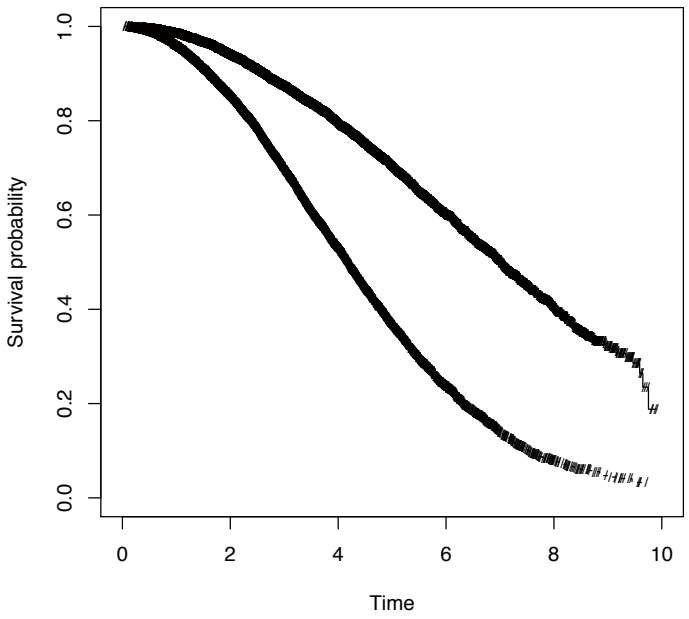
*Difference in survival proportion at 6 years

- Notice that the results do not change with the censoring distribution.

Evaluating the common choices for functional and contrast

Truth 2: Proportional hazards

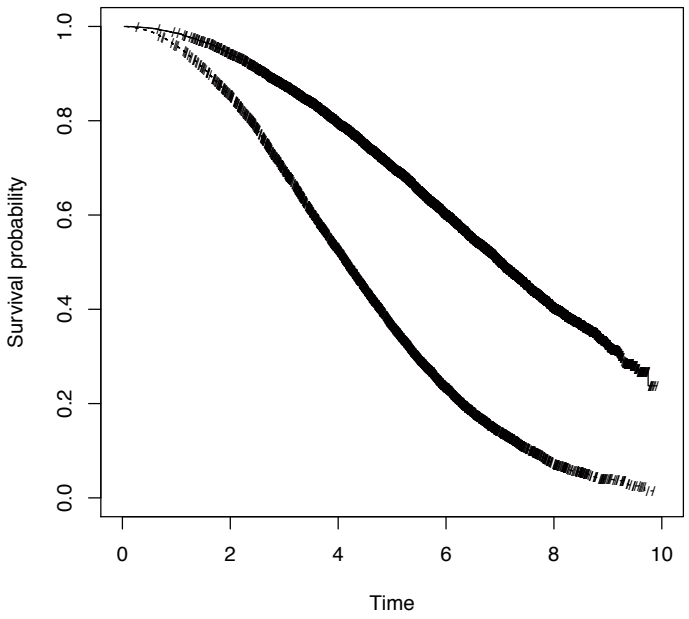
Simulation scenario: Early Censoring



Evaluating the common choices for functional and contrast

Truth 2: Proportional hazards

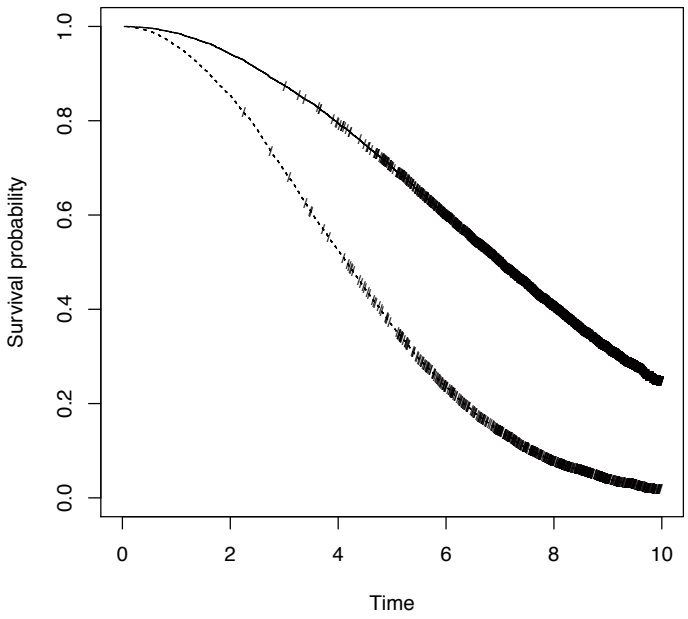
Simulation scenario: Mid Censoring



Evaluating the common choices for functional and contrast

Truth 2: Proportional hazards

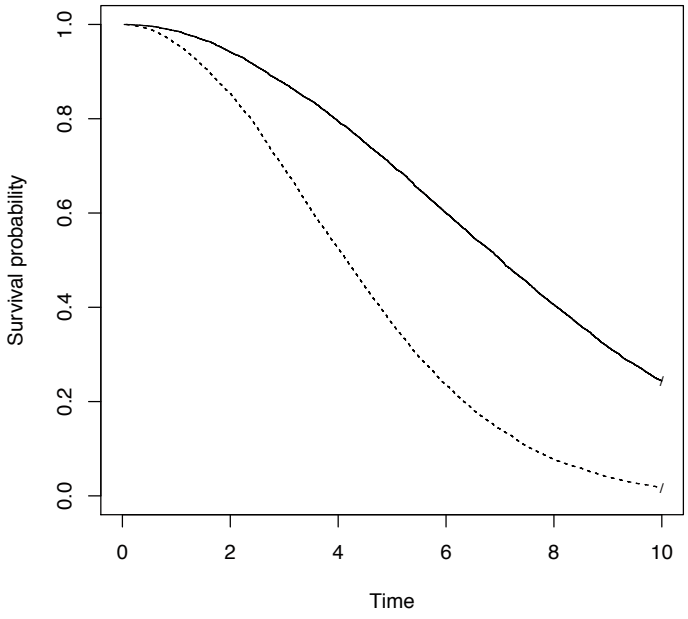
Simulation scenario: Late Censoring



Evaluating the common choices for functional and contrast

Truth 2: Proportional hazards

Simulation scenario: Censoring at 10 only (“No censoring")



Evaluating the common choices for functional and contrast

Truth 2: Proportional hazards

- Point estimates (Proportional Hazards simulations):

Scenario	Hazard Ratio	Rate Ratio	Restricted Mean	6-year Survival*
Early censoring	0.357	0.436	2.314	0.365
Mid censoring	0.351	0.441	2.366	0.368
Late censoring	0.355	0.476	2.351	0.364
No censoring	0.354	0.502	2.349	0.364

*Difference in survival proportion at 6 years

- Notice: Censoring distribution affects the RR, but not HR, RM or IT.

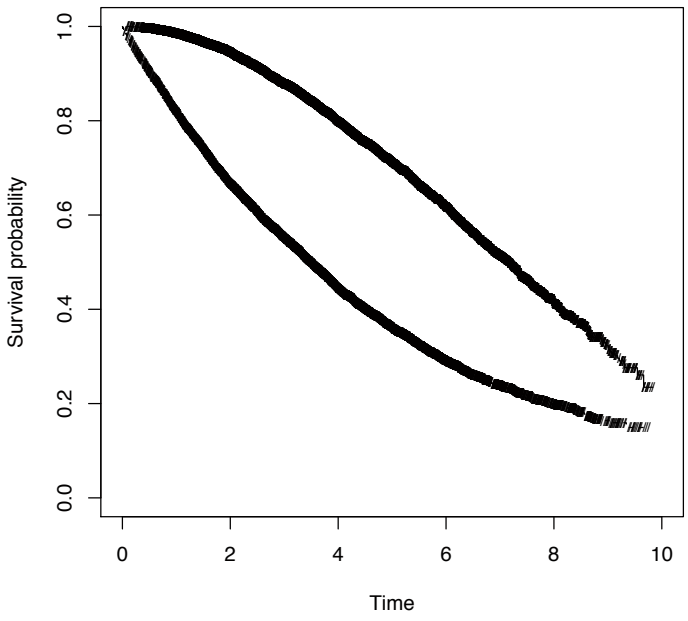
Evaluating the common choices for functional and contrast

SISCR
UW - 2016

Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline

Truth 3: Non-proportional hazards

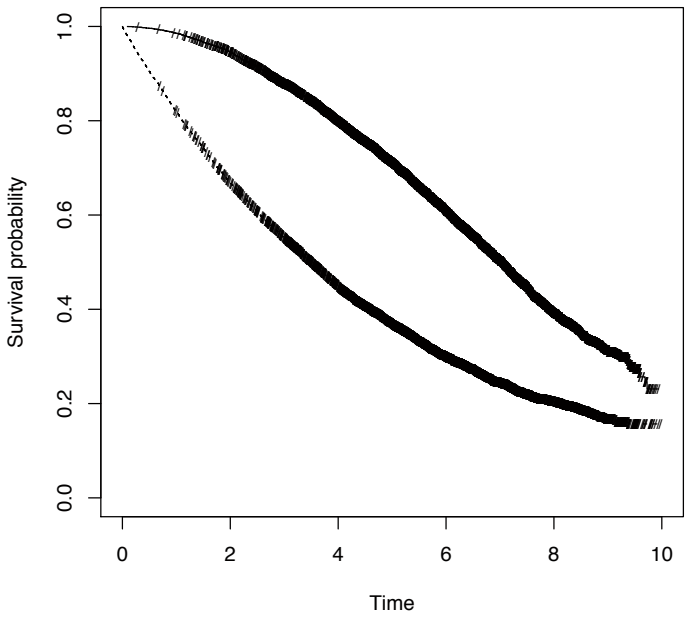
Simulation scenario: Early Censoring



Evaluating the common choices for functional and contrast

Truth 3: Non-proportional hazards

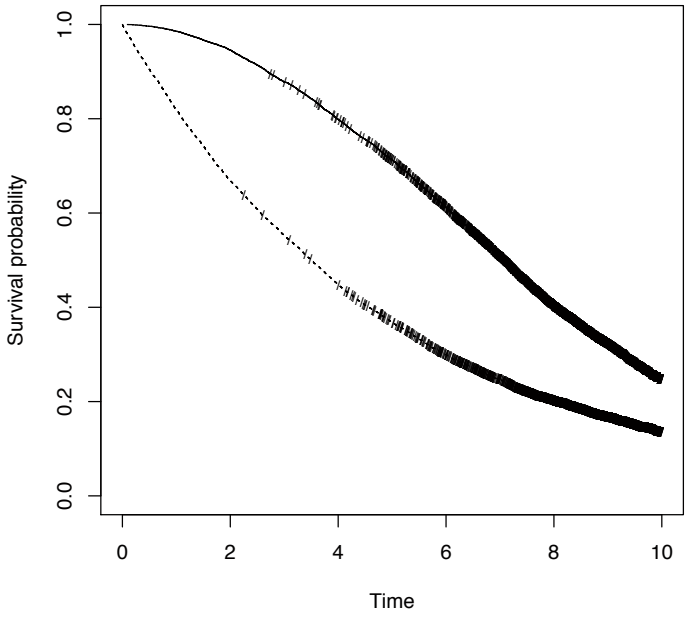
Simulation scenario:: Mid Censoring



Evaluating the common choices for functional and contrast

Truth 3: Non-proportional hazards

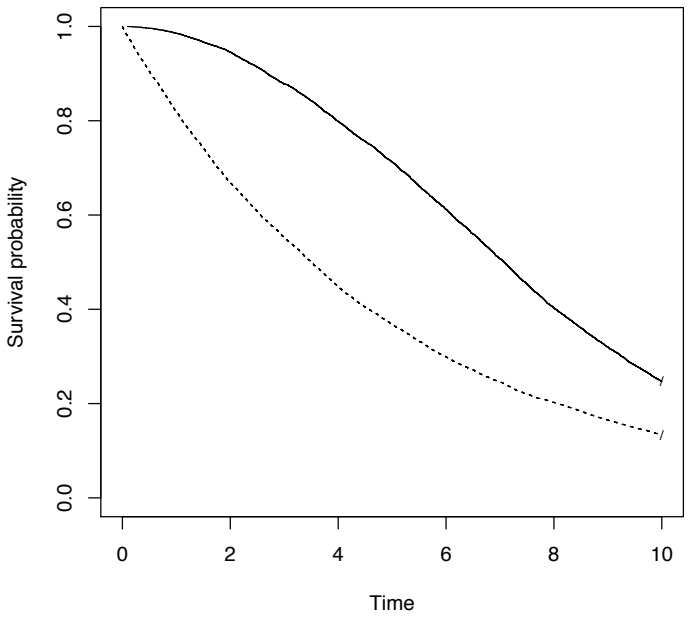
Simulation scenario: Late Censoring



Evaluating the common choices for functional and contrast

Truth 3: Non-proportional hazards

Simulation scenario: Censoring at 10 only (“No censoring")



Evaluating the common choices for functional and contrast

Truth 3: Non-proportional hazards

- Point estimates (Non-proportional Hazards simulations):

Scenario	Hazard Ratio	Rate Ratio	Restricted Mean	6-year Survival*
Early censoring	0.340	0.357	2.430	0.326
Mid censoring	0.411	0.434	2.420	0.309
Late censoring	0.483	0.512	2.467	0.313
No censoring	0.520	0.554	2.468	0.313

*Difference in survival proportion at 6 years

- Notice: Censoring distribution affects the HR and RR, but not RM or IT.

Evaluating the common choices for functional and contrast

Implications of these results

- Initial questions:
  - Which statistical models gives generalizable inference if:
    - We do not know the form of the true probability distribution(s)?
    - We do not know how treatment will affect the true distribution?
  - What are the symptoms of an answer that is not generalizable?
- Conclusions:
  - For fully parametric and semi-parametric models, inference is not consistent (i.e., it depends on the censoring distribution) unless the assumed model is true. Specifically:
    - RR only works if the number of events follows a Poisson probability distribution.
    - HR only works if there are proportional hazards.
  - Restricted mean survival does not require model assumptions and should be considered for robust inference.
  - Index time does not require assumptions, but may suffer from lack of scientific relevance and/or statistical power.

Evaluating the common choices for functional and contrast

Concluding remarks

- Standard approaches to time-to-event data:
  - (most common) Hazard ratio
  - (somewhat common) Index time
  - (rarely) Rate ratio (Poisson probability distribution)
  - (almost never) Restricted mean survival
- You should be aware that the choice of the probability model, functional, and contrast may not assure reproducible inference.
  - Changing the follow-up time may give different answer.
  - Changing the censoring distribution (early vs late) may give a different answer).
  - Therefore, your endpoint is also defined by the follow-up time and amount of follow-up.

Properties of censored time-to-event data

Competing risks

- A key assumption with time-to-event analysis is that the types of events which are censored must be “non-informative" about the event being analyzed.
- There are many potential causes of death - they are all “competing" to see which will get you first.
  - Non-informative censoring:
    - The subjects who are censored must look just like a random sample of the subjects who are still at risk. They can be neither more nor less likely to have an event in the near future than subjects who are not censored.
    - Censoring of subjects cannot be related to the risk of impending death (event). That is, subjects cannot be censored either because they are at high risk of death or because they are at low risk of death.

Properties of censored time-to-event data

Competing risks

- Example: Smoking as risk factor for cancer death.
  - Possible censoring mechanisms.
    - Subject still alive at time of data analysis.
    - Subject lost to follow-up at some point during study.
    - Subject hit by meteor.
    - Subject hit by bus.
    - Subject died of MI.
    - Subject died of emphysema.
  - Evaluation: (Might the censoring mechanism be informative about the time of the event):
    - Non-informative: Alive at time of analysis; hit by meteor; hit by bus (unless suicide); lost to follow-up (?).
    - Possibly informative: death from MI or emphysema.

Properties of censored time-to-event data

Competing risks

- Problem: there is no way to find out whether death from other causes is informative censoring.
  - It is impossible to observe two death times for the same subject.
  - Example: cannot tell when the person who died of an MI would have died of lung cancer; thus, we cannot estimate if censoring due to MI would be informative for lung cancer death.

Properties of censored time-to-event data

Competing risks

- Consequences:
  - Although it is theoretically possible to estimate the cause-specific hazard (or survival) in the presence of informative competing risks, those estimates will not generalize to changes in the distribution of competing risks. Thus, we cannot estimate what survival will be like after intervention ot decrease cause-specific mortality.
  - Potential for harm:
    - Informative competing risks can make a bad treatment look good. E.g., we can “cure" cancer by causing heart attacks in people who are most likely to die from cancer.

Properties of censored time-to-event data

Competing risks

Example: Suppose that we want to evaluate a new drug for preventing MI (fatal or non-fatal), but there might be competing risks from other causes of death.

- Analysis 1: Censor all deaths.
  - Appropriate if other deaths are non-informative.
  - Efficient (powerful) if valid
  - Irrelevant (possibly dangerous) inference with informative censoring.
- Analysis 2: Model the mechanism that leads to informative censoring:
  - Will always be based on untestable assumptions.

Properties of censored time-to-event data

SISCR
UW - 2016

Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline

Competing risks

- Analysis 3: Event-free survival (model time to MI or death from all causes, whichever comes first).
  - Minimal effect if the incidence of competing risk is low.
  - Behaves like analysis 1 if everyone has MI before competing risk.
  - Protects against false cures (e.g., preventing MI by causing death from suicide).
  - If the competing risk is non-informative, then there is some loss of power; e.g.,
    - MI: 20% on treatment; 30% on control.
    - Other causes: 30% on both arms (independent of MI)
    - MI or death: 44% on treatment, 51% on control.
- Analysis 4: Analyze survival only.
  - Ignores nonfatal MI entirely.
  - Survival is the bottom line, but it may take too long.

Properties of censored time-to-event data

Competing risks

Comments/opinions:

- Models that incorporate the censoring mechanism will be based on untestable assumptions. Wrong assumptions give wrong answers.
- Arguing against event-free survival because it requires a larger sample size ignores the potential for biased (incorrect) conclusions. In general it is more important to protect against incorrect conclusions.
- A lot of clinically important questions cannot be assessed in an analysis that looks only at survival (analysis 4):
  - What if we really just want to prevent MI's?
  - What if we want to treat a non-fatal condition?
  - What if we want to improve quality of life?
- At times it is relevant to examine cause-specific survival.

Other important designs

Change from baseline outcomes

Outline:

1. Motivation and data structure
2. Approaches to defining outcomes when an endpoint is measured at baseline and follow-up
3. Other applications
4. Statistical design (sample size and CI evaluations)

Change from baseline outcomes

Motivation and data structure

- Why measure change?
  - Within-subject change often clinically relevant
  - Usually: within-individual change is less variable.
- For example, consider the CHEST data:

Subj ID	Trt Grp	6 Minute walk distance Baseline	12-weeks	Change
1	1	167	145	-22
2	0	233	244	11
3	0	325	267	-58
4	1	214	309	95
5	1	441	457	16
6	1	447	441	-6
7	1	443	466	23
8	1	378	421	43
9	1	298	268	-30
10	0	381	316	-65
11	0	431	547	116
12	1	332	413	81
13	0	372	371	-1
14	0	300	278	-22
15	1	412	475	63
16	0	444	230	-214
17	1	215	375	160
18	1	330	410	80
19	1	300	305	5
20	1	365	360	-5

Change from baseline outcomes

Motivation and data structure

- CHEST Trial summary statistics

Study Visit	Placebo Mean (sd)	Riociguat Mean (sd)
Baseline	356.0 (74.7)	342.3 (81.9)
16-weeks	350.4 (122.2)	381.2 (119.2)
Change	-5.5 (84.3)	38.9 (79.3)

- Approaches to analysis:
  - Compare 6MWD after 16-weeks
  - Compare 16-week improvement in 6MWD
  - Linear regression: 16-week 6MWD conditional on baseline walk distance.

Change from baseline outcomes

Approaches to outcome definition

1. Evaluate difference at last time measurement time:
   - Outcome: Final measurement on each subject.
   - Functional: θ_{1F} and θ_{0F} denote the mean outcome with active and control therapy at the final measurement time.
   - Contrast: $\theta = \theta_{1F} - \theta_{0F}$.
- Example (CHEST) T-test of 16-week walk distance:

$$\begin{aligned} \hat{\theta} &= \hat{\theta}_1 - \hat{\theta}_0 = 381.20 - 350.43 = 30.765 \\ se(\hat{\theta}) &= \sqrt{\frac{119.2^2}{173} + \frac{122.2^2}{88}} = 15.87 \\ 95\%CI &= 30.765 \pm 1.9739 se = (-0.5533, 62.08) \\ p - value &= 0.055 \end{aligned}$$

- Result:
  - At 16-week mean walk distance with Riociguat is 30.77 meters farther than placebo (95% CI: -0.55 to 63.08 meters; p = 0.055).
  - Notice inconclusive result.

Change from baseline outcomes

Approaches to outcome definition

2. Evaluate change over follow-up period:
   - Outcome: Change in outcome (final minus baseline)
   - Functional:
     $\theta_1 = \theta_{1F} - \theta_{1B}$ (mean change with active treatment)
     $\theta_0 = \theta_{0F} - \theta_{0B}$ (mean change with control treatment)
   - Contrast: $\theta = \theta_1 - \theta_0$.
- CHEST: T-test of change in walk distance over 16 weeks:

$$\begin{aligned} \hat{\theta} &= \hat{\theta}_1 - \hat{\theta}_0 = 38.9 - (-5.5) = 44.41 \\ se(\hat{\theta}) &= \sqrt{\frac{79.27^2}{173} + \frac{84.32^2}{88}} = 10.82 \\ 95\%CI &= 44.41 \pm 1.9744 se = (23.05, 65.78) \\ p - value &= 6.358 \times 10^{-5} \end{aligned}$$

- Result:
  - Over 16-weeks Riociguat treatment improves mean walk distance by 44.41 meters (95% CI: 23.05 to 65.78 meters; p < 0.0001) more than the improvement with placebo.

Change from baseline outcomes

Approaches to outcome definition

3. Last measurement time adjusting for baseline:
   - Outcome: Outcome at the last measurement time (adjusting for baseline value)
   - Functional: outcome at last measurement time
   - Contrast: Difference in mean outcomes adjusted for baseline levels (θ in the following regression model):

$$\theta_{kF} = \beta_0 + \theta Tx + \beta_1 \theta_{kB}$$

   where Tx is the indicator for active treatment.

- CHEST: Fit the linear regression model:

$$Y_{iF} = \beta_0 + \theta Tx_i + \beta_1 Y_{iB}$$

```
                 Estimate Std. Error t value   Pr(>|t|)
(Intercept)      -49.2743  23.992698 -2.0537 4.1011e-02
exmpl[, "RioTx"]  46.0904  10.583863  4.3548 1.9227e-05
exmpl[, "base6"]   1.1229   0.062937 17.8417 6.2500e-47
```


Change from baseline outcomes

Approaches to outcome definition

3. Last measurement time adjusting for baseline: (cont'd)

- CHEST Result:
  - *Point estimate:* Among two populations with the same baseline walk distance, after 16 weeks a population taking riociguat will end up walking 46.1 meters farther than a population taking placebo.
  - *95% CI:* (25.35, 66.84)
  - *P-value:* 1.922×10^{-5}.

Change from baseline outcomes

Approaches to outcome definition

- Comparison of approaches:

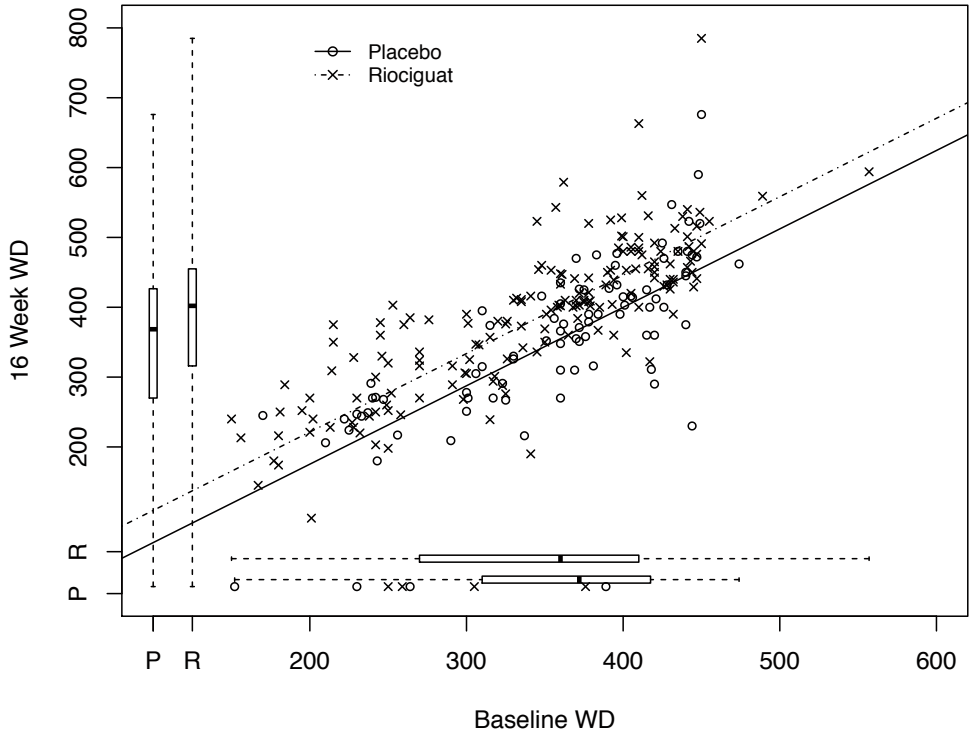
Approach	Estimate	95% CI	p-value
16-week difference	30.77	(-0.55, 62.08)	0.055
Change from baseline	44.41	(23.05, 65.78)	6.358×10^{-5}
Adjusting for baseline	46.1	(25.35, 66.84)	1.922×10^{-5}

Change from baseline outcomes

Graphical depiction of the CHEST results

SISCR
UW - 2016

Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline



Change from baseline outcomes

Interpretation of the plot

- Notes on previous graph:
  - Boxplots on left represent the data used for a t-test of 16-week differences.
  - Regression lines are the result of the above linear regression analysis.
    - Vertical distance between regression lines is 46.1 meters (i.e., the effect of riociguat adjusted for baseline walk distance).
  - Bloxplots on bottom show no confounding (similar distribution of baseline WD).
- This example shows how we can increase power by adjusting for a precision variable.
  - I now illustrate the general behavior in a series of graphs

Change from baseline outcomes

Precision variables (in linear regression)

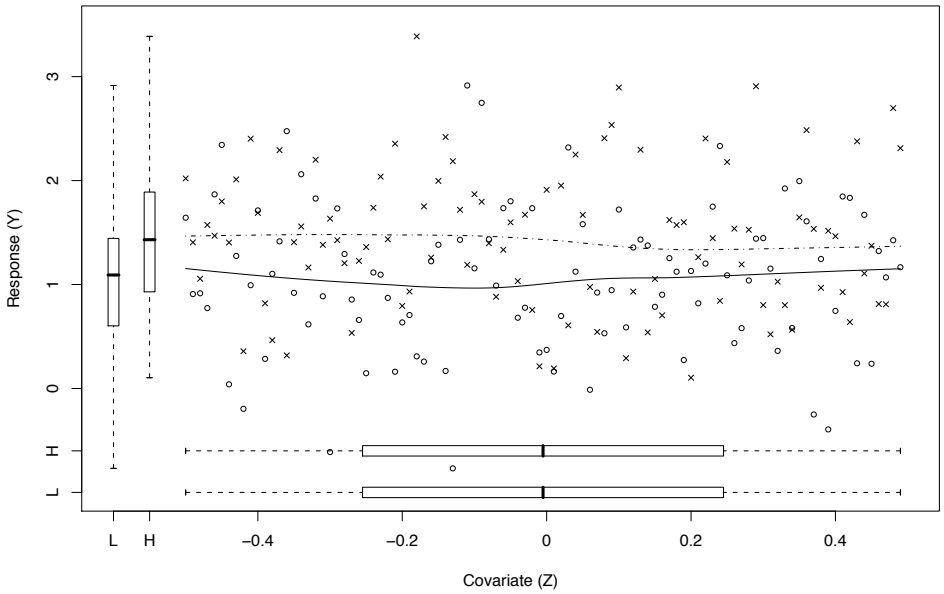
- Recall: A precision variable reduces “noise" (extraneous variation) so that the relationship between outcome and the primary explanatory variable is more precise.
- A precision variables must be:
  - unrelated to the primary explanatory variable.
  - an independent predictor of outcome.
- Adjusting for a precision variable increases precision for the comparison of interest.

Change from baseline outcomes

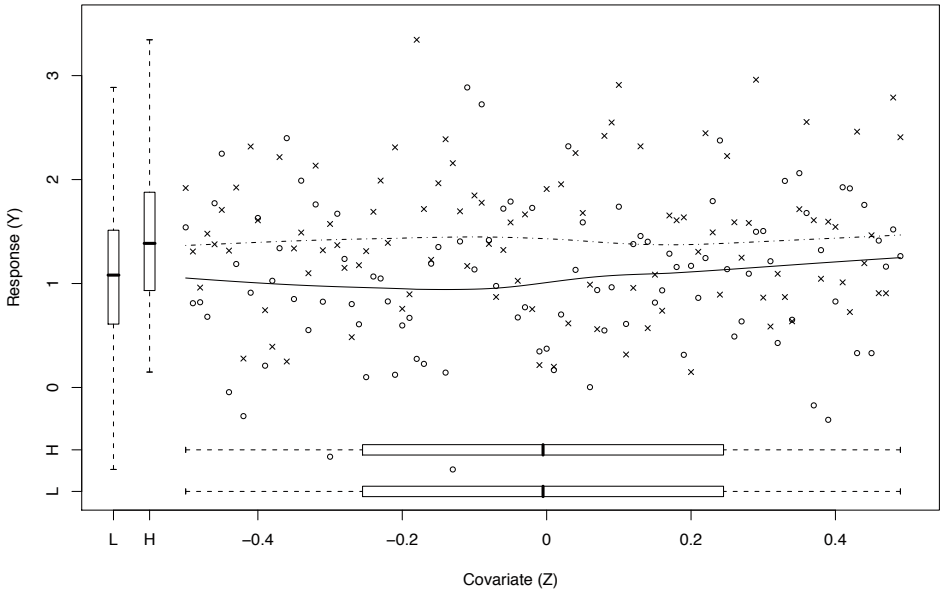
Example (Precision variable)

- The nature of a precision variable can be illustrated using scatterplots. Let:
  - Y denotes outcome
  - X denotes primary explanatory variable (2-categories: H and L)
  - Z denotes a covariate (precision variable)
- We are interested in the relationship between X and Y.
- In the following plots:
  - The relationship between X and Y is fixed.
  - There is no relationship between X and Z (the precision variable is unrelated to the explanatory variable).
  - The relationship (correlation) between Y and Z is increasing.

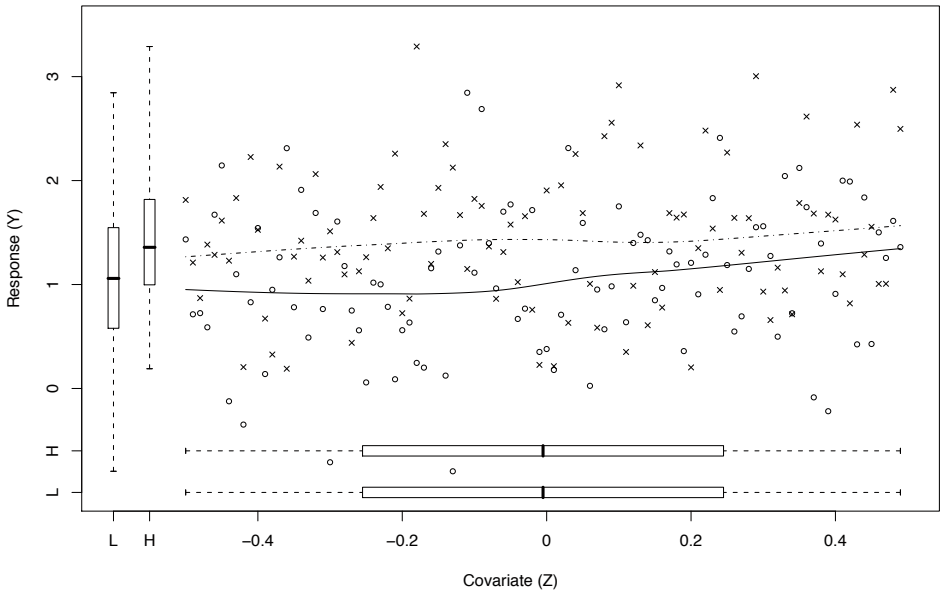
Example (Figure 1a)



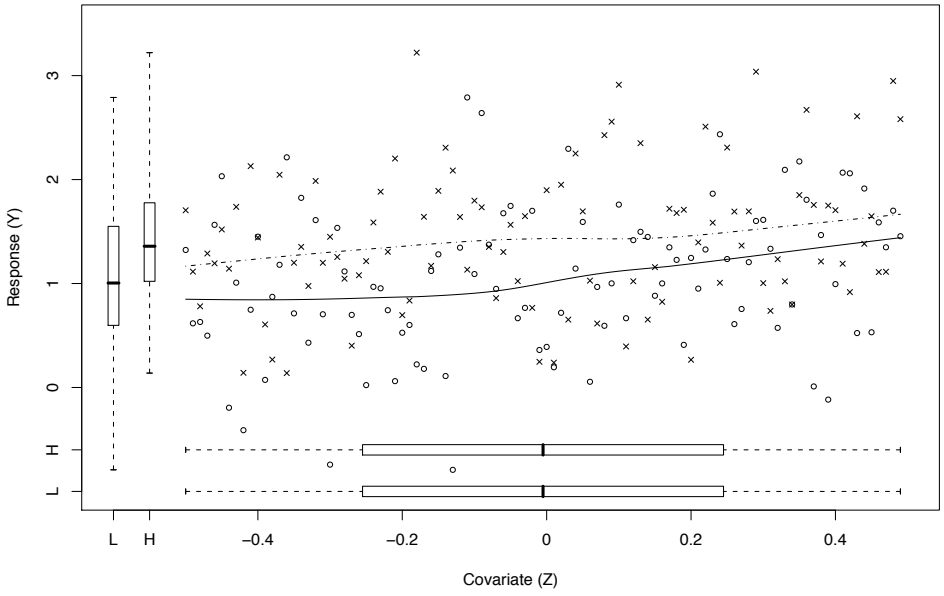
Example (Figure 1b)



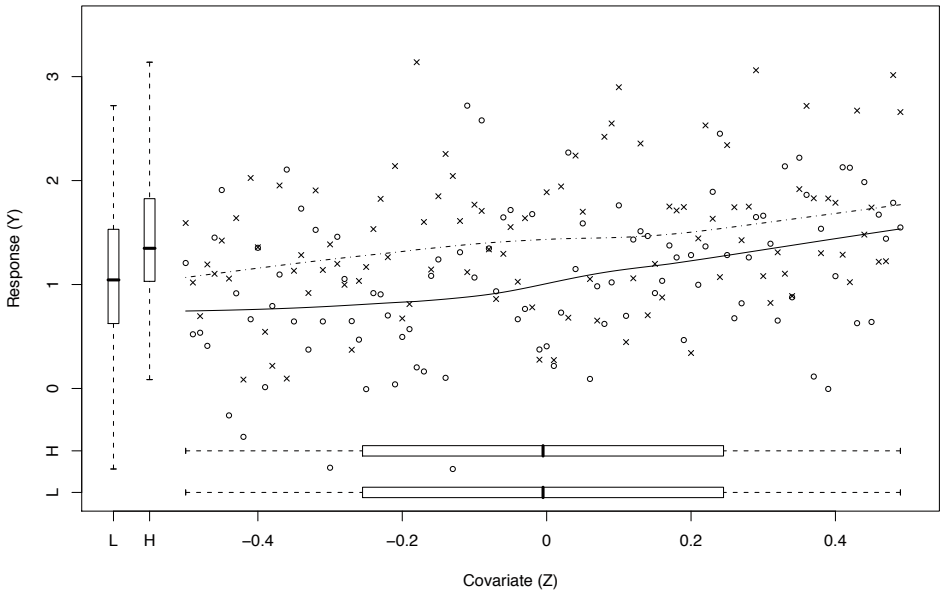
Example (Figure 1c)



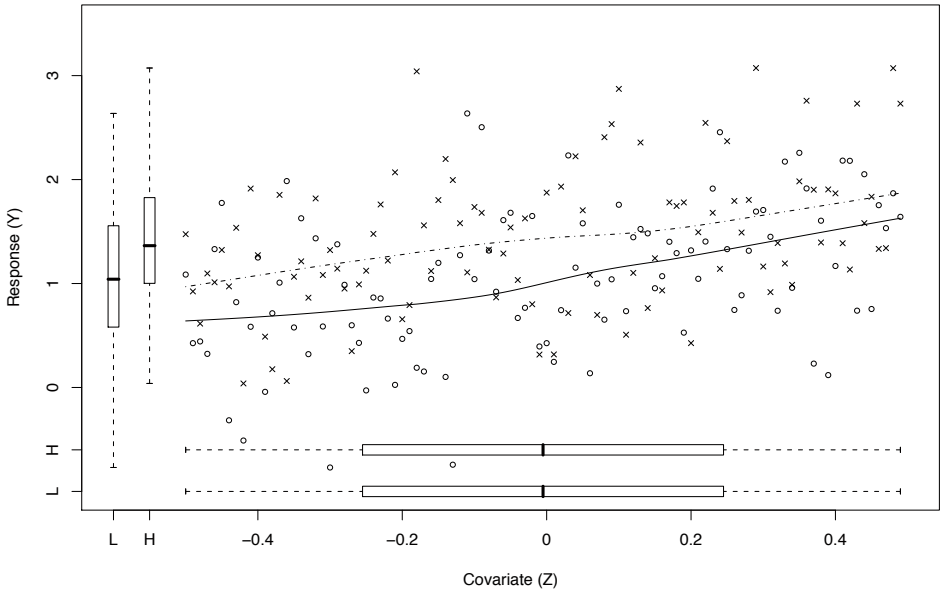
Example (Figure 1d)



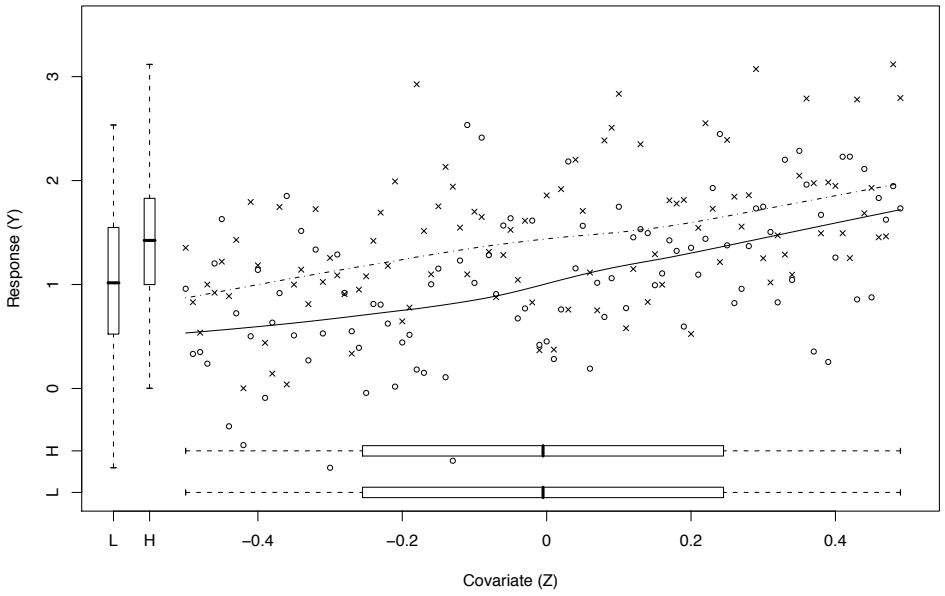
Example (Figure 1e)



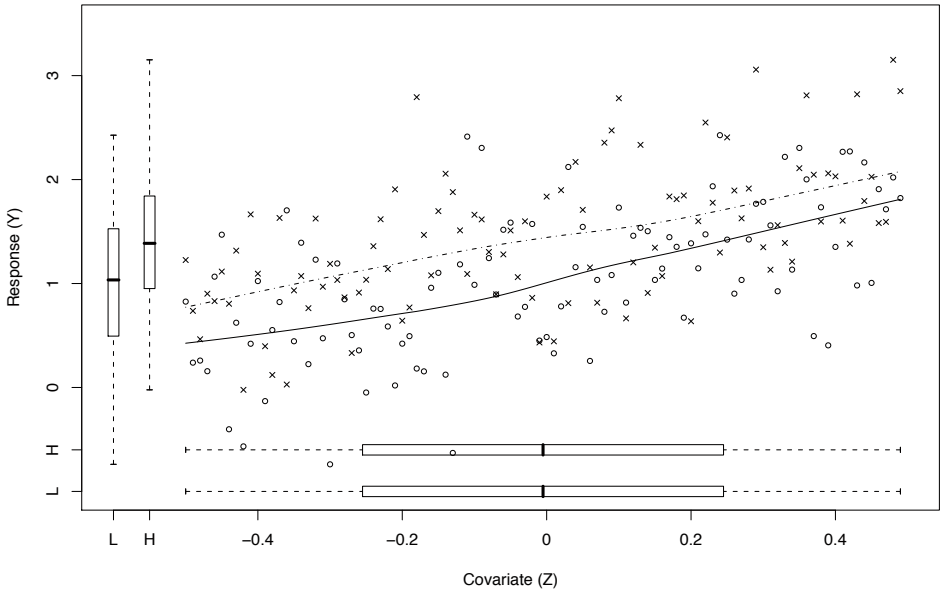
Example (Figure 1f)



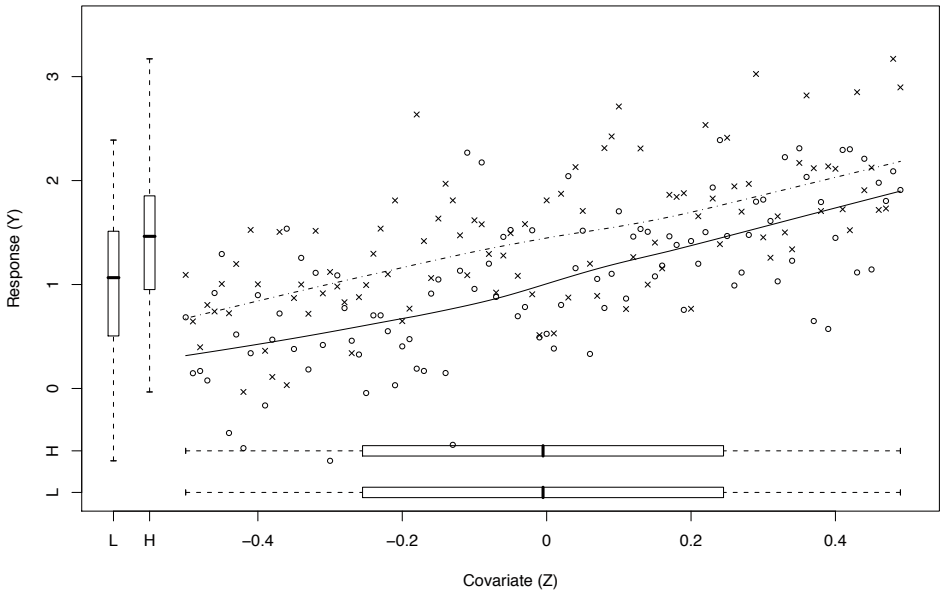
Example (Figure 1g)



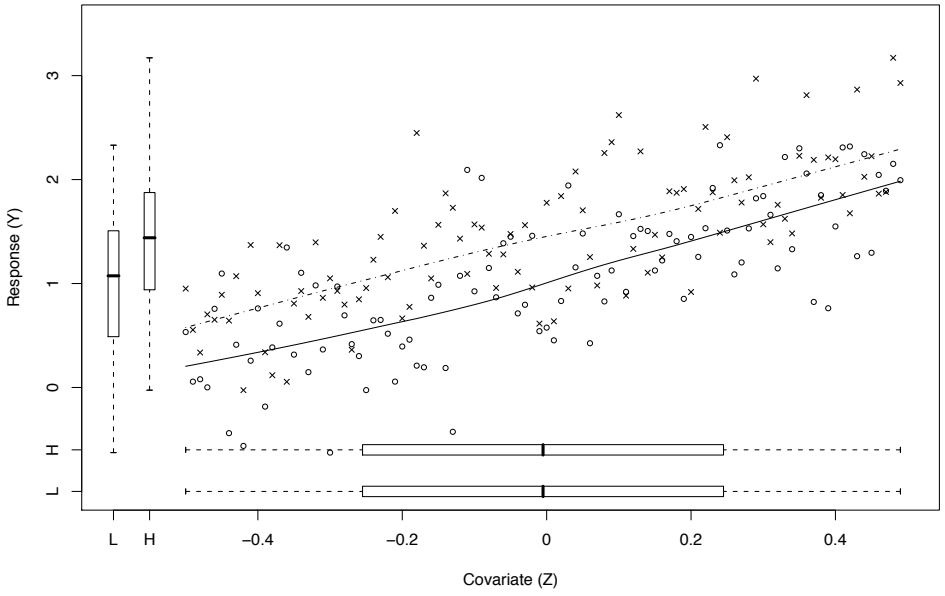
Example (Figure 1h)



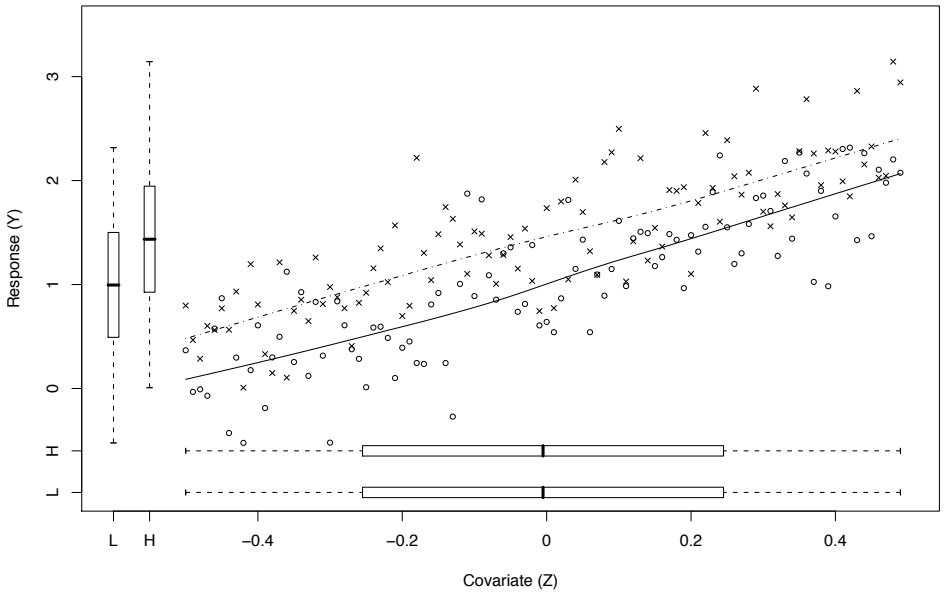
Example (Figure 1i)



Example (Figure 1j)



Example (Figure 1k)



Change from baseline outcomes

Example (Precision variable)

- We would like to compare the results of an analysis if we ignore the precision variable (the “crude" difference) with the results after we adjust for the precision variable (the “adjusted difference").

	(Z,Y)	Crude Difference		Adjusted Difference	
	Correlation	Estimate	SE	Estimate	SE
Fig 1a	0.000	0.389	0.098	0.389	0.099
Fig 1b	0.082	0.390	0.098	0.390	0.098
Fig 1c	0.164	0.391	0.098	0.391	0.097
Fig 1d	0.246	0.393	0.098	0.393	0.096
Fig 1e	0.328	0.395	0.098	0.395	0.093
Fig 1f	0.410	0.399	0.098	0.399	0.090
Fig 1g	0.492	0.404	0.098	0.404	0.086
Fig 1h	0.574	0.409	0.098	0.409	0.081
Fig 1i	0.656	0.416	0.098	0.416	0.074
Fig 1j	0.739	0.425	0.098	0.425	0.066
Fig 1k	0.821	0.437	0.099	0.437	0.056

Change from baseline outcomes

SISCR
UW - 2016

Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline

Example (Precision variable)

Notice:

- Crude and adjustments are identical
- Standard error (SE) of the adjusted estimate is smaller than the standard error of the crude estimate

  (Note: smaller SE gives more power)
- The precision of the adjusted estimate increases with the correlation between Y and Z
- The precision variable is “explaining" some of the variation (reducing the noise) in the primary comparison

Change from baseline outcomes

Other applications

- Other situations in which the primary analysis is adjusted for baseline values:
  - Common to adjust for stratification variables
  - May adjust for scientific interpretability
  - May adjust for comparability to previous studies
- It is important to pre-specify all adjustments as part of your primary analysis.

Change from baseline outcomes

Implications for power and information

- The above examples and graphical illustration illustrate general principles:
  - Variance (precision) of various approaches to defining outcomes with change from baseline data.
  - It is particularly clear when $\sigma_0 = \sigma_1 \ (= \sigma)$ and $N_1 = N_0 \ (= N)$:

Variance when comparing only the last time point:

$$var(\hat{\theta}) = \frac{2\sigma^2}{N} \quad (1)$$

Variance when comparing change from baseline:

$$var(\hat{\theta}) = \frac{4\sigma^2(1-\rho)}{N} \quad (2)$$

Variance when comparisons are adjusted for baseline:

$$var(\hat{\theta}) = \frac{2\sigma^2(1-\rho^2)}{N} \quad (3)$$

- In all of the above, ρ is the correlation between baseline and follow-up measures.

Change from baseline outcomes

Implications for power and information

- The above relationships can be used to prove:
  - If $\rho < 0.5$, then it is more powerful to analyze follow-up differences
    (i.e., DO NOT compare change from baseline).
  - If $\rho > 0.5$ then it is more powerful to analyze change from baseline than follow-up differences.
  - It is always more powerful to us regression to adjust for baseline:
    - This is also known as “Analysis of covariance" (ANCOVA).
    - The CHEST paper refers to it as the “least-squares" estimate.

Change from baseline outcomes

Implications for power and information

- Relative sample size of the analytic approaches:
  - Change from baseline relative to follow-up only:

$$\frac{4\sigma^2(1-\rho)}{2\sigma^2} = 2(1-\rho)$$

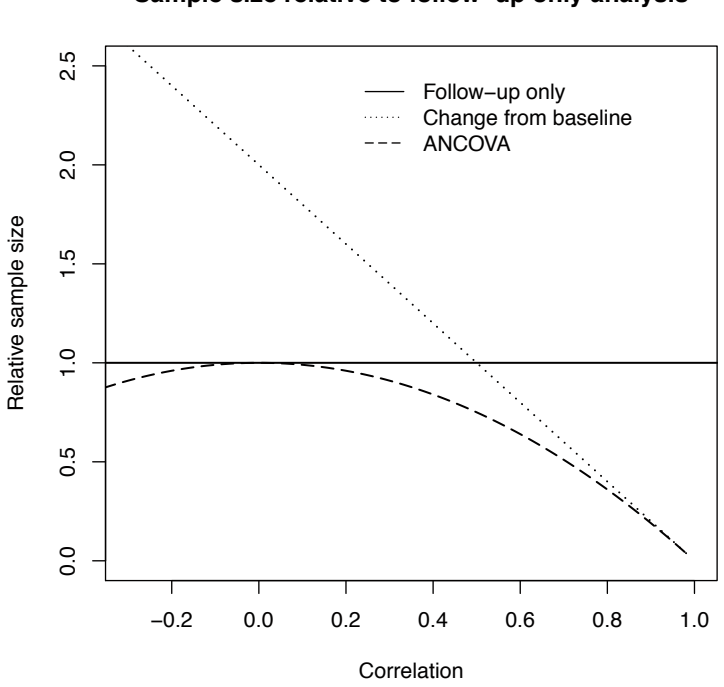
  - ANCOVA relative to follow-up only:

$$\frac{2\sigma^2(1-\rho^2)}{2\sigma^2} = (1-\rho^2)$$

Change from baseline outcomes

Implications for power and information

Relative sample size of analytic approaches as function of correlation:



Change from baseline outcomes

SISCR
UW - 2016

Implications for power and information

Frequent asked questions (FAQ) about the ANCOVA analysis:

- The ANCOVA model described above fits parallel lines.
  - What happens if the lines are not parallel?
    * Non-parallel lines represents interaction; treatment works better (or worse) for low baseline values.
    * Interactions are explored in subsequent trials.
  - What happens if the relationships are not linear?
    * Not a problem as long as baseline distribution is the same in both treatment groups (assured by randomization).
    * The line represents the first order approximation to the curve (i.e., is it treading up, down, or flat?).

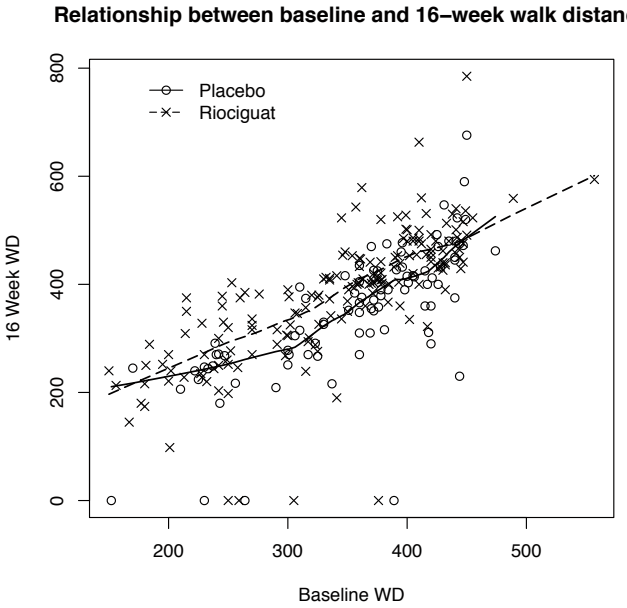
Time-to-event outcomes
Properties of time-to-event data
Parameterizing time-to-event outcomes
Competing risks
Change from baseline

Change from baseline outcomes

Implications for power and information

Example: ANCOVA FAQ's in the CHEST trial

- No evidence for substantial non-linearity:

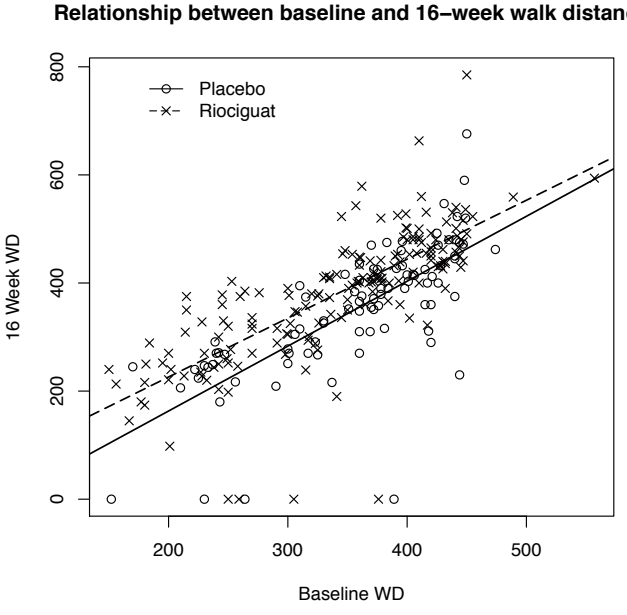


Change from baseline outcomes

Implications for power and information

Example: ANCOVA FAQ's in the CHEST trial

- Separate lines in each treatment group are nearly parallel:



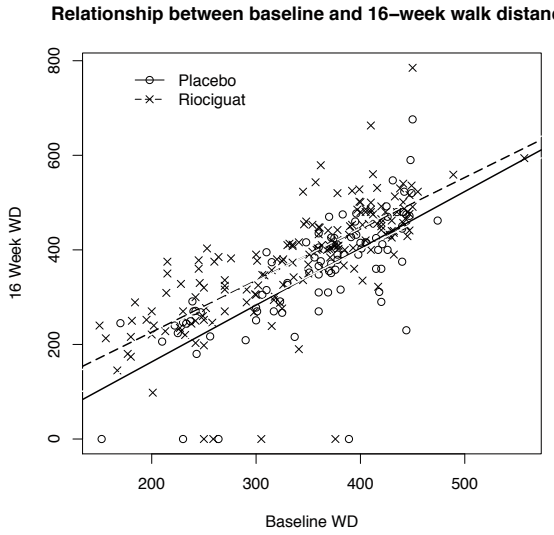
- No problem with interaction.

Change from baseline outcomes

Implications for power and information

Example: ANCOVA FAQ's in the CHEST trial

- Separate lines in each treatment group are nearly parallel:



- No problem with interaction.